

# An Introduction to `mpitbR`

Girela, Ignacio

1 Feb, 2025

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Multidimensional Poverty Measurement</b>	<b>3</b>
2.1	The Alkire-Foster method step-by-step . . . . .	3
2.2	Decomposition properties . . . . .	5
2.2.1	Incidence and intensity of poverty . . . . .	5
2.2.2	Indicators breakdown . . . . .	5
2.2.3	Subgroup Decomposition . . . . .	6
2.3	The global MPI . . . . .	6
2.3.1	Components . . . . .	6
2.3.2	Unit of identification and unit of analysis . . . . .	7
2.3.3	Main data sources . . . . .	8
<b>3</b>	<b>The <code>mpitbR</code> package</b>	<b>8</b>
3.1	Installation . . . . .	8
3.2	Data processing . . . . .	8
3.2.1	Load package and data . . . . .	9
3.2.2	Missing values . . . . .	9
3.2.3	Household survey design . . . . .	10
3.3	Define the multidimensional poverty measurement project . . . . .	10
3.4	Cross-sectional estimates . . . . .	11
3.4.1	A quick start . . . . .	11
3.4.2	Analyzing poverty across population groups . . . . .	13
3.4.3	Indicator-specific measures analysis . . . . .	15
	<b>References</b>	<b>20</b>

# 1 Introduction

`mpitbR` is a package for calculating Alkire-Foster class measures of multidimensional poverty. The measurement method proposed by Alkire and Foster (2011) distinguishes itself for its versatility in adjusting the indicators, weighting schemes, and poverty cut-offs to different contexts. Indeed, this method is the formal scaffold of the global Multidimensional Poverty Index (MPI) (Alkire and Santos 2014), a comparably international measure of acute poverty, yearly published by the Oxford Poverty and Human Development Initiative (OPHI) and the United Nations Development Programme (UNDP). In addition, other regional and national MPIs have been created by adapting the global MPI to better address local realities.

The global MPI is presented for more than 100 countries, together with ten constituent indicators aligned with Sustainable Development Goals (SDGs), as well as with recommendations of the World Bank’s *Atkinson Commission on Monitoring Global Poverty*. Committed to transparency and collaboration, OPHI publishes all technical files to reproduce their findings. This includes all the Stata do-files to prepare the microdata for generating the global MPI indicators. Then, the global MPI estimates are calculated using the ‘mpitb’ Stata package developed by Nicolai Suppa (Suppa 2023).

The `mpitbR` package faithfully replicates the estimation procedures of the original Stata ‘mpitb’ package, ensuring methodological consistency for researchers using different programming languages. By offering an R implementation, `mpitbR` contributes to a more integrated and collaborative research ecosystem around the global MPI, aligning with OPHI’s encouragement of international collaboration.

This vignette describes basic usage guide for this package, illustrated with real world examples. First, we begin with an introduction to the Alkire-Foster method and the global MPI. Subsequently, we demonstrate how to install and start using the `mpitbR` package. For those already familiar to the AF method, can directly proceed to Section 3 and explore multidimensional poverty analysis in practice. This section introduces good practices and caveats in data processing for MPI calculations, along with steps for computing AF measures for a single year. We also provide code for plotting results, which can be valuable for personal multidimensional poverty research projects. This vignette aims to complement the `mpitbR` package documentation. For further details on function usage, please refer to the reference manual.

## 2 Multidimensional Poverty Measurement

### 2.1 The Alkire-Foster method step-by-step

Due to the widely acknowledgement of the multidimensional nature poverty both in academic and policy circles, this century has witnessed an significant emergence of multidimensional poverty measurement methodologies. Among these, the ‘dual cut-off’ framework proposed by Alkire and Foster (2011) has gained prominent attention for its flexibility and key-policy properties.

The Alkire-Foster (AF) method can be summarized in the following steps (for a detailed explanation, see Alkire et al. 2015):

#### 1. Establish the data source

One of the most salient features of the AF measure is the ability to consider the multiple deprivations faced by the poor jointly. Therefore, all the information ought to come from the same data source, commonly household surveys.

When designing a multidimensional poverty measure, stakeholders decide which data source will best align with the poverty measure. As we will see, this selection is linked to two following steps.

#### 2. Determine the unit of analysis

Depending on the purpose of the MPI in question, the unit of analysis will be defined, i.e., who or what is being studied (individuals, households or even communities). This step influences the choice of indicators, the data source, and interpretation of results. For simplicity, we will refer to ‘person’ as the unit of analysis.

#### 3. Select the dimensions and indicators

Poverty is a complex phenomenon, however, for measuring purposes it is necessary to define which dimensions of human development the measure will focus on. Each dimension will be represented by a set of  $d \in \mathbb{N}$  indicators (e.g., years of schooling and children school attendance are the two indicators that represents education dimension in the global MPI).

To represent people’s well-being in all dimensions, an  $n \times d$  achievement matrix  $X$  is defined, where each element  $x_{ij} \in \mathbb{R}_+$  is an ordinal variable that denote the achievement or well-being status of the person  $i$  in the  $j$ -th indicator, for  $i = 1, \dots, n$  and  $j = 1, \dots, d$ .

#### 4. Define each indicator deprivation cut-off

A first cut-off  $z_j$  is defined as the minimum level of achievement necessary for being non-deprived in indicator  $j$ , i.e., if  $x_{ij} < z_j$ , person  $i$  is considered deprived in indicator  $j$ . We denote the deprivation cut-offs as  $z = (z_1, \dots, z_d)$ .

#### 5. Obtain the deprivation matrix

Then apply the deprivation cut-offs vector to each of the  $n$  observations to obtain the deprivation matrix  $g^0$ , where each element is a binary variable denoting the deprivation status of person  $i$  in indicator  $j$ :  $g_{ij}^0 = 1$  if  $x_{ij} < z_j$ , and  $g_{ij}^0 = 0$  if  $x_{ij} \geq z_j$ . In matrix form,

$$g^0 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

represents the deprivation matrix for five observations and four indicators. The deprivation vector of the first person  $g_1 = [1, 0, 1, 0]$  reflects she is deprived in the first and third indicator.

## 6. Assign weights to each indicator

The weight  $w_j$  of each dimension and indicator reflects their relative importance, where  $\sum_{j=1}^d w_j = 1$ . In practice, an equal nested weighting scheme is used: dimensions are weighted equally as well as each indicator within the dimension. The weights vector is denoted as  $w = (w_1, \dots, w_d)$ .

## 7. Calculate the deprivation score

By combining  $w$  and  $g^0$  is possible to build the weighted deprivation matrix  $\bar{g}^0$ , where the non-null values of the deprivation matrix  $g_{ij}^0$  are replaced by the corresponding value of  $w_j$ . In matrix form, this is equivalent to  $\bar{g}^0 = \text{diag}(w) \times g^0$ . Using the previous example:

$$\bar{g}^0 = \begin{bmatrix} w_1 & 0 & w_3 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & w_2 & w_3 & w_4 \\ w_1 & 0 & 0 & 0 \\ w_1 & 0 & w_3 & 0 \end{bmatrix}$$

Then, this information is aggregated by the weighted deprivations to obtain the deprivation score  $c_i$  for each person, i.e.,  $c_i = \sum_{j=1}^d w_j g_{ij}^0 = \sum_{j=1}^d \bar{g}_{ij}^0$ .

Let assume that indicators are equally weighted,  $w = (1/4, 1/4, 1/4, 1/4)$ . Then, the deprivation score vector is  $c = (1/2, 1/4, 3/4, 1/4, 1/2)$ .

## 8. Select the poverty cut-off to identify the poor

A second cut-off  $k$  is compared with the deprivation score to determine whether the person is poor or not, i.e., if  $c_i \geq k$ , the  $i$ -th person is poor. Then, the poverty cut-off  $k$  represents the minimum proportion of weighted indicators a person needs to experience to be considered multidimensional poor. This procedure consists of censoring the non-poor from the analysis.

This identification criterion allows obtaining both the censored (weighted) deprivation matrix,  $g^0(k)$ , and the censored deprivation score,  $c_i(k)$ . The former consists on censoring the non-poor from the (weighted) deprivation matrix, i.e., replace the row entries of the non-poor by a vector of zeros. Analogously, the censored deprivation score for each person is equal to  $c_i$  if the person is identified as poor, and 0 otherwise.

In our example, assume that  $k = 1/2$  (i.e., a person is considered poor if she experience deprivation in half or more of the weighted indicators). Then, the second and fourth people are not poor and the  $g^0(k)$  matrix is expressed as

$$g^0(k) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

and the censored deprivation vector is  $c = (1/2, 0, 3/4, 0, 1/2)$ .

## 9. Calculate the Multidimensional Poverty Index

The Multidimensional Poverty Index (MPI) is also referred as the *Adjusted Headcount Ratio*, and denoted as  $M_0$ . It is defined as the expected value of the censored deprivation score, i.e.,

$$M_0 = \frac{1}{n} \sum_{i=1}^n c_i(k)$$

Following the example,  $M_0 = \frac{1}{5}(1/2 + 0 + 3/4 + 0 + 1/2) = 0.350$ .

## 2.2 Decomposition properties

The MPI satisfies several desirable properties for a poverty measure. Notably, it allows for valuable decomposition analyses used in policy design. Firstly, the MPI can be disaggregated by two partial measures: incidence and intensity of poverty. These measures enhances the interpretation and understanding of the overall MPI value. Secondly, the MPI can be decomposed by individual indicators. This analysis pinpoints the specific dimensions where poverty is most prevalent, informing policymakers on the most critical areas for intervention. Thirdly, the MPI can be decomposed by population subgroups, such as gender, ethnicity, or rural/urban location. This granularity reveals disparities in poverty experiences across different segments of society, guiding targeted interventions.

### 2.2.1 Incidence and intensity of poverty

The  $M_0$  measure can be expressed as the product of two partial indices representing the incidence and intensity of poverty. Recall that people are identified as poor if  $c_i \leq k$ . Let  $q$  denote the number of people identified as poor. Then, we can multiply and divide  $M_0$  by  $q$  and rearrange some terms to compute the incidence and intensity.

$$M_0 = \frac{1}{n} \sum_{i=1}^n c_i(k) \times \frac{q}{q} = \frac{q}{n} \times \frac{1}{q} \sum_{i=1}^n c_i(k) = H \times A$$

where  $H$  and  $A$  denote the incidence and intensity of poverty, respectively.

The incidence of poverty  $H$  represents the proportion of multidimensional poor people in a society and it is represented in percentage terms. Recapitulate our previous example. In that case, the first, the third, and the fifth person were identified as poor,  $q$ , over all the three people  $n = 5$ . Then, the incidence is equal to  $H = 3/5 = 0.6$ , i.e., 60.0% of the population is multidimensional poor.

On the other hand, the intensity of poverty  $A$  represents the average weighted deprivations (score) that the poor experience. In our example,  $A = (1/2 + 0 + 3/4 + 0 + 1/2)/3 = 58.33\%$ , i.e., on average, the poor people are deprived in 58.33% of the weighted indicators.

Here is why the MPI is also named as *Adjusted Headcount Ratio*: it represents the proportion of multidimensionally poor people, adjusted by the average intensity they experience. Notably, if all the poor face deprivations across all indicators, the MPI would be equal to the incidence  $H$  of poverty. Then, if we multiply  $H$  by the intensity  $A$ ,  $M_0$  now represents the proportion of weighted deprivations that the poor experience within a society out of the total potential deprivations they could experience overall.

### 2.2.2 Indicators breakdown

A relevant property of the MPI for policy analysis is its decomposability by censored indicators. Note that in the step-by-step procedure, we construct the MPI by first aggregating information across censored indicators by columns and then across individuals by rows. The same result can be achieved by reversing this aggregation order.

To see this, take the  $M_0$  equation and rearrange the aggregation order, i.e.,

$$M_0 = \frac{1}{n} \sum_{i=1}^n c_i(k) = \frac{1}{n} \sum_{i=1}^n \left[ \sum_{j=1}^d w_j g_{ij}^0(k) \right] = \sum_{j=1}^d w_j \left[ \frac{1}{n} \sum_{i=1}^n g_{ij}^0(k) \right]$$

where, we define  $h_j(k) = \frac{1}{n} \sum_{i=1}^n g_{ij}^0(k)$  as the censored headcount ratio of indicator  $j$ . Similarly, the uncensored headcount ratio is calculated using the uncensored deprivation matrix:  $h_j = \frac{1}{n} \sum_{i=1}^n g_{ij}^0$ . The MPI can be expressed as the weighted sum of the censored headcount ratio of all indicators:

$$M_0 = \sum_{j=1}^d w_j h_j(k)$$

In our example, the uncensored indicator headcounts are  $h = (60\%, 40\%, 60\%, 20\%)$ , while the censored indicator headcounts are  $h(k) = (40\%, 20\%, 60\%, 20\%)$ . Comparing the difference between the uncensored and censored indicators yields valuable insights into the prevalence of deprivation among the poor population.

On the other hand, the absolute and percentage contribution of each indicator is reported. The absolute contribution of each indicator is determined by multiplying its weight by the censored indicator value. The percentage contribution is calculated as follows:

$$\phi_j = w_j \frac{h_j(k)}{M_0}$$

where  $\phi_j$  is the relative contribution of indicator  $j$  in the Adjusted Headcount Ratio. Following the example, the percentage contributions are 28.58%, 14.29%, 42.85%, 14.29%, which sum to 1.

### 2.2.3 Subgroup Decomposition

Another key policy-relevant property of the MPI is the subgroup decomposition. The  $M_0$ , and other associated AF measures ( $H$ ,  $A$ ,  $h_j$ ,  $h_j(k)$ ), can be calculated for various population subgroups, such as age cohorts, regions, living areas, gender, ethnicity, and educational attainment. This allows the overall measure to be expressed as the sum of the measures by group, weighted by the share of the population of that subgroup. Formally,

$$M_0 = \sum_{l=1}^L \nu_l M_0^l$$

where  $\nu_l = n_l/n$  is the proportion of people belonging to the population subgroup  $l$ , and  $M_0^l$  is the Adjusted Headcount Ratio of the population subgroup  $l$ .

In our example, consider a population divided into two regions: North and South, where first two individuals are from the North region, while the remaining individuals are from the South region. The population share of the South region is greater than the North (60% vs. 40%). The MPI for the North region is 0.250, whereas the South region exhibits a higher MPI of 0.416.

## 2.3 The global MPI

### 2.3.1 Components

We have previously introduced the global MPI, an international measure of acute multidimensional poverty, aligned with the Sustainable Development Goals (SDGs). It comprises ten deprivation indicators grouped into three poverty dimensions: two for health, two for education, and six for living standards. Their weights and deprivation cut-offs are defined as follows:

- **Health** (1/3 weight)
  - Child Mortality (1/6 weight): Deprived if any child in the household has died in the five years preceding the survey.
  - Nutrition (1/6 weight): Deprived if any child under five years old is underweight or any adult is undernourished.

- **Education** (1/3 weight)
  - Years of Schooling (1/6 weight): Deprived if no household member aged 10 years or older has completed at least six years of schooling.
  - School Attendance (1/6 weight): Deprived if any school-aged child is not currently enrolled in school.
- **Living Standards** (1/3 weight)
  - Cooking Fuel (1/18 weight): Deprived if the household relies on solid fuels (such as wood, dung, or charcoal) for cooking.
  - Sanitation (1/18 weight): Deprived if the household lacks access to improved sanitation facilities.
  - Drinking Water (1/18 weight): Deprived if the household lacks access to improved drinking water sources or if improved drinking water is more than a 30-minute round trip to collect.
  - Electricity (1/18 weight): Deprived if the household lacks access to electricity.
  - Housing (1/18 weight): Deprived if the household’s housing structure is inadequate (e.g., natural materials for walls, floors, or roofs).
  - Assets (1/18 weight): Deprived if the household does not own more than one of the following assets: radio, television, telephone, computer, animal cart, bicycle, motorbike, or refrigerator.

The weighting methodology employs nested equal weights. All dimensions are assigned equal weights, and within each dimension, all indicators are considered equally relevant. In the global MPI, each of the three dimensions (Health, Education, and Living Standards) receives a weight of one third.

Within the Health and Education dimensions, each indicator receives an equal weight of 1/2. Since the dimension carries a weight of 1/3, each indicator within that dimension has an overall weight of 1/6.

Analogously, each indicator of the Living Standards dimension receives a 1/6 weight, resulting in an overall weight of 1/18 for each indicator within that dimension.

The global MPI establishes a poverty cut-off of  $k = 1/3$  to measure acute poverty. A person is considered multidimensionally poor if they are deprived in 33.33% or more of the weighted indicators. Alternatively, given the weight assigned to each dimension, experiencing deprivation in at least one dimension can also be interpreted as being poor. Furthermore, the global MPI reports results using poverty cut-offs of 20% and 50% to denote vulnerability and severe poverty status, respectively.

Finally, the annual global MPI report includes other key information such as the indicator breakdown and percentage contributions (including uncensored levels of deprivation in each indicator), disaggregated details by certain population subgroups (rural-urban, age groups, and subnational regions), and other key estimates for inference (standard errors and confidence intervals).

### 2.3.2 Unit of identification and unit of analysis

When employing the MPI, a critical distinction lies between the unit of identification and the unit of analysis. These concepts significantly influence how poverty is measured and understood within the MPI framework, with substantial implications for poverty analysis

The unit of identification is the entity from which data is collected for poverty assessment. This could be an individual, a household, or even a community. The choice of unit carries crucial assumptions. For instance, selecting the household as the unit implies that poverty affects all members of the household equally. Since it uses available information on all household members, the global MPI utilizes the household as the unit of identification.

The unit of analysis refers to the level at which data is aggregated and analyzed to draw conclusions about poverty. While the household might be the unit of identification, the global MPI analyzes data at

the individual level using appropriate sampling weights. Reporting results by individuals facilitates the exploration of gendered or age-related disparities and enables the examination of intra-household variations in poverty experience.

In conclusion, the careful consideration of both the unit of identification and the unit of analysis is paramount for robust and insightful poverty assessments using the MPI.

### 2.3.3 Main data sources

The MPI uses information from three main sources that are publicly available and consistent for most developing countries. These sources are:

- The Demographic and Health Surveys (DHS)
- The Multiple Indicators Cluster Survey (MICS)
- National Household Surveys: If information from the former surveys is not available for a specific country, the MPI may use data from other surveys conducted within that country, as long as those surveys cover the same topics. For instance, data from the national survey ‘Encuesta de Condiciones de Vida 2013-2014’ was used to calculate the global MPI in Ecuador.

Every year, the global MPI report, country briefings, data tables and technical files are updated. In some cases, indicators definitions are refined. All particular national variations are documented in the methodological notes for the year in which the MPI was released. The major revision was in 2018 to align the indicators with SDGs (Alkire et al. 2022).

## 3 The `mpitbR` package

Having established the fundamental concepts in the previous section, we now delve into the core of this vignette, focusing on the measurement of multidimensional poverty. We will do so while carefully considering best practices and relevant considerations.

### 3.1 Installation

The simplest way to install `mpitbR` is to download and install it directly from CRAN by typing the following command in R console:

```
install.packages("mpitbR")
```

Another way is to install the development version from the `mpitbR` Github repository:

```
library(devtools)
install_github("girelaignacio/mpitbR")
```

### 3.2 Data processing

This vignette utilizes a preprocessed Benin DHS surveys (DHS06) included in this package. These datasets were prepared using Stata do-files to process the raw microdata according to the specifications outlined by the OPHI and the UNDP for their annual global MPI reports.



These datasets include two unique identifiers (household and individual IDs), variables related to the survey design (primary sampling unit, stratum, and sampling weight), demographic characteristics (sex, living area, and region), and ten columns representing each of the ten indicators used in the global MPI calculation. The columns `d_cm`, `d_nutr`, `d_satt`, `d_educ`, `d_elct`, `d_wtr`, `d_sani`, `d_hsg`, `d_ckfl`, `d_asst` represents Child Mortality, Nutrition, School Attendance, Years of Schooling, Electricity, Water, Sanitation, Housing, Cooking Fuel, and Assets indicators, respectively.

### 3.2.1 Load package and data

Below, we load the installed `mpitbR` package and present a few rows of the preprocessed Benin 2006 round DHS microdata, `'ben_dhs06'`:

```
library(mpitbR)

head(ben_dhs06)

##   hh_id  ind_id strata psu   weight   sex area region d_cm d_nutr d_satt
## 1 10001 1000101     1   1 0.503804  male urban Alibori   1     0     1
## 2 10001 1000102     1   1 0.503804 female urban Alibori   1     0     1
## 3 10001 1000103     1   1 0.503804  male urban Alibori   1     0     1
## 4 10001 1000104     1   1 0.503804  male urban Alibori   1     0     1
## 5 10001 1000105     1   1 0.503804 female urban Alibori   1     0     1
## 6 10001 1000106     1   1 0.503804  male urban Alibori   1     0     1
##   d_educ d_elct d_wtr d_sani d_hsg d_ckfl d_asst
## 1     1     1     1     1     1     1     0
## 2     1     1     1     1     1     1     0
## 3     1     1     1     1     1     1     0
## 4     1     1     1     1     1     1     0
## 5     1     1     1     1     1     1     0
## 6     1     1     1     1     1     1     0
```

! Note that the `mpitbR` package operates directly on the uncensored deprivation matrix, where all the indicators columns contain binary values (0 and 1). Consequently, the package cannot be used for earlier stages of the analysis, such as generating deprivation indicators or making normative decisions regarding their definition.

### 3.2.2 Missing values

Multidimensional poverty indicators often contain missing values due to non-response from household members. Missing data are more common in cohort-specific deprivations, where the unit of analysis is the individual. For example, when the unit of analysis is the household, if a child's anthropometric measurements are missing, not only is the child assigned a missing value, but it also impacts the poverty status of the entire household.

To ensure accurate MPI calculations, an important step involves verifying that all missing values within indicators are consistently assigned to all members of the unit of identification (e.g., the household). On the other hand, data cleaning should focus on retaining only the relevant columns, including those containing survey design data (e.g., primary sampling units (PSUs), weights, strata), and variables related to population groups, as it in `ben_dhs06` data.

The code below explore the total number of missing values for all the indicators columns in `ben_dhs06` data using `tidyverse` package.

```

# Load `tidyverse` package
library(tidyverse)

# Count missing values by all the deprivation indicators columns
# (all their names start with d_*)
indicators_NAs <- ben_dhs06 %>%
  summarise(across(grep("^d_", colnames(ben_dhs06)), ~sum(is.na(.))))

print(indicators_NAs)

```

```

##   d_cm d_nutr d_satt d_educ d_elct d_wtr d_sani d_hsg d_ckfl d_asst
## 1 1882  5650   234     9     89  101    75   41   372    0

```

```

# Now compare the total number of missing values in the dataset with the indicators.
total_NAs <- sum(is.na(ben_dhs06))
print(total_NAs == sum(indicators_NAs))

```

```
## [1] TRUE
```

We observe a higher frequency of missing values within the Health dimension indicators compared to other dimensions. Given that these missing values are exclusively associated with deprivation indicators, we can employ the `na.omit()` function to directly remove observations containing any missing data.

```
ben_dhs06 <- na.omit(ben_dhs06)
```

! If one observation (unit of analysis) has a missing value, all other observations belonging to the same group (unit of identification) should exhibit a missing value for that variable. While OPHI do-files prevent this to occur, practitioners should remain attentive to such inconsistencies in their measurement projects.

### 3.2.3 Household survey design

Household surveys, the primary data source for multidimensional poverty measurement, employ complex survey designs. In order to ensure the reliability of point estimates and their associated standard errors and confidence intervals, crucial for statistical inference, `mpitbR` accounts for complex survey designs by utilizing methods from the `survey` R package (Lumley 2024). This is another reason why it is important to remove missing values, as they can introduce subtle and potentially difficult-to-detect biases into the estimates generated by `survey` package functions.

We now define the survey design using the `svydesign` function from the `survey` package, considering the primary sampling units (`psu`), sampling weights (`weight`), and strata (`strata`) information in the data

```

# Load `survey` library
library(survey)

# Define the survey design
svydata <- svydesign(ids = ~psu, weights = ~weight, strata = ~strata, data = ben_dhs06)

```

## 3.3 Define the multidimensional poverty measurement project

Once the survey design is set, we specify the MPI measurement project settings. This includes defining our data source (`svydata` object in this case), identifying the dimensions and assigning indicator columns to each

dimension, and optionally providing a label for our project (a brief name with a short description). To do this, we utilize the `mpitb.set` function.

Since we are reproducing the global MPI for Benin DHS 2006, we group our indicator columns in health, education and living standards dimensions using a list.

```
# Group indicators by dimension
indicators <- list(hl = c("d_nutr", "d_cm"),
                  ed = c("d_satt", "d_educ"),
                  ls = c("d_elct", "d_sani", "d_wtr", "d_hsg", "d_ckfl", "d_asst"))
```

Next, pass the data and indicators as arguments to the `mpitb.set` function.

```
# Set the multidimensional poverty project
set <- mpitb.set(data = svydata, indicators = indicators,
                 name = "ben_dhs06", desc = "Benin global MPI 2006")
```

`set` is an object of class `mpitb_set` that contains all the relevant information of the MPI measurement project for further use.

### 3.4 Cross-sectional estimates

The core of this package is `mpitb.est`, designed for estimating the MPI and their partial measures (intensity, incidence, indicator-specific measures by population subgroups).

The `mpitb.est` function offers several arguments to allow for customized MPI calculations. Here below, we outline the key arguments (for a comprehensive list of arguments and their descriptions, type `?mpitb.est` in your R console):

- `set` is the multidimensional poverty measurement project settings previously defined with `mpitb.set`.
- `k` is the vector of poverty cut-offs (values between 1 and 100).
- `weights` is a vector specifying the weighting scheme for each dimension. By default, equal nested weights are used, as employed in the global MPI.
- `measures` refers to the main aggregate measures to be calculated ( $M_0$ ,  $H$ , or  $A$ ).
- `indmeasures` are all the indicator-specific measures, such as censored and raw headcount ratios, and their absolute and percentage contribution to overall poverty.

By default, all measures are estimated.

#### 3.4.1 A quick start

As an example, we will now demonstrate the estimation of the global MPI for Benin in 2006.

```
# Estimate the Benin global MPI 2006
estimate.01 <- mpitb.est(set, k = 33, measures = "M0", indmeasures = NULL)
```

```

##          ***** SPECIFICATION *****
## Call:
## mpitb.est.mpitb_set(set = set, klist = 33, measures = "M0", indmeasures = NULL)
## Name:  ben_dhs06
## Weighting scheme:  equal
## Description:  Benin global MPI 2006
## -----
##
## Dimension 1:  hl 0.333                (d_nutr, d_cm)
## Dimension 2:  ed 0.333                (d_satt, d_educ)
## Dimension 3:  ls 0.333 (d_elct, d_sani, d_wtr, d_hsg, d_ckfl, d_asst)
## -----
##
## Indicator 1:  d_nutr 0.1667
## Indicator 2:  d_cm 0.1667
## Indicator 3:  d_satt 0.1667
## Indicator 4:  d_educ 0.1667
## Indicator 5:  d_elct 0.0556
## Indicator 6:  d_sani 0.0556
## Indicator 7:  d_wtr 0.0556
## Indicator 8:  d_hsg 0.0556
## Indicator 9:  d_ckfl 0.0556
## Indicator 10: d_asst 0.0556
##
##          ***** ESTIMATION *****
## -----
## Partial AF measures: ' M0 ' under estimation... DONE
##
##          ***** RESULTS *****
## -----
## Parameters
## Subgroups:  1
## Poverty cut-offs (k):  1
##
## *Notes:
##   Confidence level: 95 %
##   Parallel estimations:  FALSE

```

Upon execution, the `mpitb.est` function displays a message that includes the function call itself, a list of dimensions with their assigned indicators and corresponding weights (allowing users to verify the setup), the specific measures that have been estimated, relevant estimation parameters such as the number of poverty cut-offs and subgroups analyzed, and other important features like the confidence level used for calculating confidence intervals and whether parallel estimation has been employed. This detailed message can be suppressed by setting the `verbose` argument to `FALSE` within the `mpitb.est` function call.

The estimation results are stored in the `estimate.01` object, which is an instance of the `mpitb_est` class. This object is a list containing two data frames: `lframe`, which encompasses all cross-sectional estimates for each level of analysis, and `cotframe`, which contains measures of change over time for each level of analysis. However, since this example does not involve changes-over-time analysis, the `cotframe` element will be `NULL`. This structure allows for flexible storage and retrieval of both cross-sectional and longitudinal poverty estimates.

```

# Take a glance at the results
as.data.frame(estimate.01$lframe)

```

```
##           b           se           ll           ul subg loa measure ctype k indicator
## 1 0.4381315 0.00604726 0.426297 0.4500369  nat nat      M0  lev 33      NA
```

This displays the raw data frame of our estimates. The first four columns provide the point estimate (`b`), the standard error of the point estimate (`se`), which accounts for the survey design, and the lower and upper bound of the confidence interval (`lb` and `ub`), with confidence intervals calculated considering the measures as proportions. `measure` and `k` columns indicate the specific AF measure and poverty cut-off, respectively, to which each estimate corresponds.

### 3.4.2 Analyzing poverty across population groups

We have previously established that the MPI can be calculated as the sum of the MPIs of mutually exclusive population subgroups, weighted by the population share of each group. This decomposition is crucial for identifying sociodemographic disparities in poverty distribution. Typically, global MPI analyses explore differences across rural-urban areas, different age cohorts, and by gender.

To calculate the MPI for specific subgroups, utilize the `over` argument within the `mpitb.est` function. This argument accepts a character vector specifying the column names of the population groups for which poverty analysis is desired.

Let's examine disparities between rural and urban areas in Benin during 2006. We can execute the following code:

```
# Include living areas in the MPI calculation
estimate.02 <- mpitb.est(set, k = 33, measures = "M0", indmeasures = NULL,
                        over = "area", verbose = FALSE)

# View results
as.data.frame(estimate.02$lframe)
```

```
##           b           se           ll           ul subg loa measure ctype k
## 1 0.4381315 0.006047260 0.4262970 0.4500369  nat nat      M0  lev 33
## 2 0.5313261 0.007133483 0.5173004 0.5453025 rural area      M0  lev 33
## 3 0.2852113 0.010101845 0.2657999 0.3054505 urban area      M0  lev 33
## indicator
## 1      NA
## 2      NA
## 3      NA
```

Again, results are presented as a data frame. The population group (in this case, `area`) is identified in `loa` column, which stands for 'level of analysis'. The corresponding MPI estimates for each level of analysis are labeled by each subgroup in the `subg` column. Note that another `loa` entry exists with a unique subgroup called "nat". This represents the MPI estimate calculated across all the observations within the data set, generally at the national level. To exclude the overall national estimate, set the `overall` argument to `FALSE` within the `mpitb.est` function.

The following code generates a barplot using `ggplot2` to visualize and compare national, and living-area poverty levels.

```
# Save complete subgroups names to be used in the plots
subg_names <- c("National", "Rural", "Urban")

plt_data.MPI <- as.data.frame(estimate.02$lframe) %>%
  # Replace the subgroup names by their complete names
```

```

mutate(subg = factor(stringi::stri_replace_all_regex(
  subg, pattern = c("nat","rural","urban"),
  replacement = subg_names, vectorize = F), levels = subg_names))

# Plot!
plt <- ggplot(plt_data.MPI,
  aes(x = subg, y = b, fill = subg)) +
geom_bar(stat = "identity", position = "dodge", width = 0.5) +
# Add confidence intervals
geom_errorbar(aes(ymin = ll, ymax = ul), width = 0.15) +
# Axis labels
labs(x = "Level of Analysis", y = "MPI", fill = "Subgroups") +
# White background
theme_bw() +
# Legend position
theme(legend.position = "bottom") +
# Bars color
scale_fill_manual(values = c("#8C8C8CFF", "#88BDE6FF", "#FBB258FF"))

# Show plot
plt

```

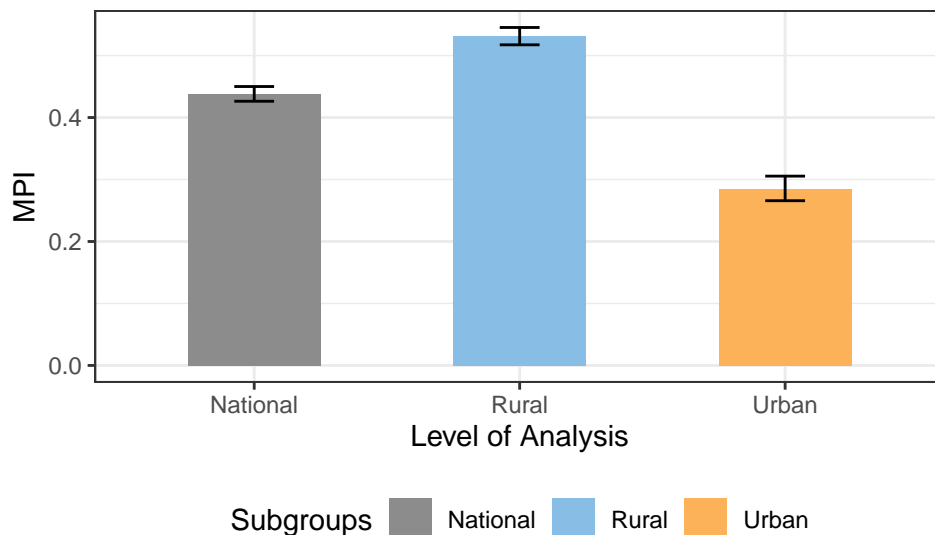


Figure 1: Multidimensional Poverty by Living Areas in Benin 2006

Figure 1 demonstrates that poverty levels in rural areas are significantly higher than in urban areas of Benin (0.531 and 0.285, respectively). Confidence intervals are displayed at the top of each bar. A good practice for comparing poverty levels is to examine if the confidence intervals do not overlap. In this example, it is possible to infer that multidimensional poverty in rural Benin is statistically higher than in urban Benin.

To further investigate whether higher multidimensional poverty in rural Benin is primarily driven by a larger proportion of poor individuals (incidence,  $H$ ) or by the rural poor experiencing greater deprivation (intensity,  $A$ ), we can incorporate these two measures into our previous analysis using the `measures` argument.

```

# Include incidence H and intensity A in the MPI calculation
estimate.03 <- mpitb.est(set, k = 33, measures = c("MO","H","A"), indmeasures = NULL,
  over = c("area"), verbose = FALSE)

```

```
# Explore coefficients of H and A
# Incidence
coef(subset(estimate.03$lframe, measure == "H" & loa == "area"))
```

```
## Subgroup Level of analysis Cut-off Coefficient
## 1 rural area 33 0.876
## 2 urban area 33 0.527
```

```
# Intensity
coef(subset(estimate.03$lframe, measure == "A" & loa == "area"))
```

```
## Subgroup Level of analysis Cut-off Coefficient
## 1 rural area 33 0.607
## 2 urban area 33 0.541
```

While the intensity of poverty is statistically higher in rural areas of Benin than in urban areas, the difference in intensity compared to urban areas is not as pronounced as the difference in incidence (87.59% of the rural population experiencing multidimensional poverty compared to 52.73% in urban areas).

! A final important caveat: when analyzing multidimensional poverty across different subgroups, avoid subsetting the dataset to isolate specific groups before specifying the measurement project with `mpitb.set` function. Subsetting the data can impact the degrees of freedom, potentially compromising the accuracy of statistical inferences.

### 3.4.3 Indicator-specific measures analysis

Breaking down the MPI by indicators provides valuable insights in multidimensional poverty analysis. We can compare the censored and uncensored headcount ratios of each indicator and explore their contribution (absolute and percentage) to the MPI. This granular analysis can be further refined by examining these contributions across different population subgroups, offering a high-resolution lens of poverty within a society.

As mentioned earlier, the argument `indmeasures` in the `mpitb.est` function encompasses all the indicator-specific measures and, by default, all of them are calculated. This is why we previously set this argument to `NULL`.

```
# Estimate indicator-specific measures
# We specify nothing in `indmeasures`
# since all of them are calculated by default.
estimate.04 <- mpitb.est(set, k = 33, measures = "M0", over = "area")
```

```
## ***** SPECIFICATION *****
## Call:
## mpitb.est.mpitb_set(set = set, klist = 33, measures = "M0", over = "area")
## Name: ben_dhs06
## Weighting scheme: equal
## Description: Benin global MPI 2006
## -----
##
## Dimension 1: hl 0.333 (d_nutr, d_cm)
## Dimension 2: ed 0.333 (d_satt, d_educ)
## Dimension 3: ls 0.333 (d_elct, d_sani, d_wtr, d_hsg, d_ckfl, d_asst)
```

```

## -----
##
## Indicator 1:  d_nutr 0.1667
## Indicator 2:   d_cm 0.1667
## Indicator 3:  d_satt 0.1667
## Indicator 4:  d_educ 0.1667
## Indicator 5:  d_elct 0.0556
## Indicator 6:  d_sani 0.0556
## Indicator 7:   d_wtr 0.0556
## Indicator 8:   d_hsg 0.0556
## Indicator 9:  d_ckfl 0.0556
## Indicator 10: d_asst 0.0556
##
##          ***** ESTIMATION *****
## -----
## Partial AF measures: ' MO ' under estimation... DONE
##
## -----
## Indicator-specific measures: ' hd hdk actb pctb ' under estimation... DONE
##
##          ***** RESULTS *****
## -----
## Parameters
## Subgroups: 2
## Poverty cut-offs (k): 1
##
## *Notes:
##   Confidence level: 95 %
##   Parallel estimations: FALSE

# View results
head(estimate.04$lframe)

```

```

##           b           se           ll           ul  subg  loa  measure  ctype  k
## 1  0.4381315 0.006047260 0.4262970 0.4500369  nat  nat      MO  lev 33
## 2  0.5313261 0.007133483 0.5173004 0.5453025 rural area  MO  lev 33
## 3  0.2852113 0.010101845 0.2657999 0.3054505 urban area  MO  lev 33
## 113 0.4527050 0.007121543 0.4387647 0.4667199  nat  nat      hd  lev NA
## 212 0.5065955 0.009153211 0.4886248 0.5245491 rural area  hd  lev NA
## 311 0.3642776 0.011294546 0.3424047 0.3867260 urban area  hd  lev NA
## indicator
## 1      <NA>
## 2      <NA>
## 3      <NA>
## 113  d_nutr
## 212  d_nutr
## 311  d_nutr

```

The `indicator` column now makes more sense. It clearly indicates to which indicator the estimated measure (measure column) corresponds.

In practice, a valuable exercise involves comparing visually the uncensored and censored indicators' headcount ratios and the contributions of each indicator within each population subgroup. This deeper exploration reveals how the structure of poverty changes across different segments of society.



```

# Save complete indicators names to be used in the plots
indicators_names <- c("Nutrition", "Child Mortality",
                     "School Attendance", "Years of Schooling",
                     "Electricity", "Sanitation", "Water",
                     "Housing", "Cooking Fuel", "Assets")

# Rearrange data to create fancier plots :)
plt_data.hd <- as.data.frame(estimate.04$lframe) %>%
# Filter by the indicators headcount ratios
filter(measure == "hd" | measure == "hdk") %>%
# Replace the indicators names by their complete names
mutate(indicator = factor(stringi::stri_replace_all_regex(
  indicator, pattern = unlist(indicators),
  replacement = indicators_names, vectorize = F), levels = indicators_names)) %>%
# Replace the subgroup names by their complete names
mutate(subg = factor(stringi::stri_replace_all_regex(
  subg, pattern = c("nat", "rural", "urban"),
  replacement = subg_names, vectorize = F), levels = subg_names)) %>%
# Replace the measure abbreviation by their complete names
mutate(measure = ifelse(measure == 'hd', 'Uncensored',
                       ifelse(measure == 'hdk', 'Censored', measure)))

# Plot!
plt <- ggplot(plt_data.hd,
  aes(x = indicator, y = b, fill = measure)) +
  geom_bar(stat = "identity", width = 0.5,
  position=position_dodge()) +
  # Headcount as percentage
  scale_y_continuous(labels = scales::percent) +
  # Legend position
  theme(legend.position = "right") +
  # Axis Labels
  labs(y = "Indicators Headcount ratios", fill = "Measure", x = "Indicators") +
  # White background
  theme_bw() +
  # facet by population subgroups
  facet_grid(rows = vars(subg)) +
  # Fit indicators names by rotating them
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Show plot
plt

```

Figure 2 illustrates the raw and censored headcount ratios for each indicator in Benin. The raw headcount represents the overall percentage of the population deprived in a specific indicator, while the censored headcount focuses on the percentage of the multidimensionally poor population experiencing that particular deprivation.

The figure reveals several key findings. Firstly, living standards indicators generally exhibit higher deprivation rates compared to other indicators. Secondly, the population in Benin demonstrates significant deprivation in nutrition and years of schooling.

Furthermore, remarkable disparities exist between rural and urban areas. In rural regions, the raw and censored headcount ratios for most indicators tend to be closer, indicating a strong correlation between

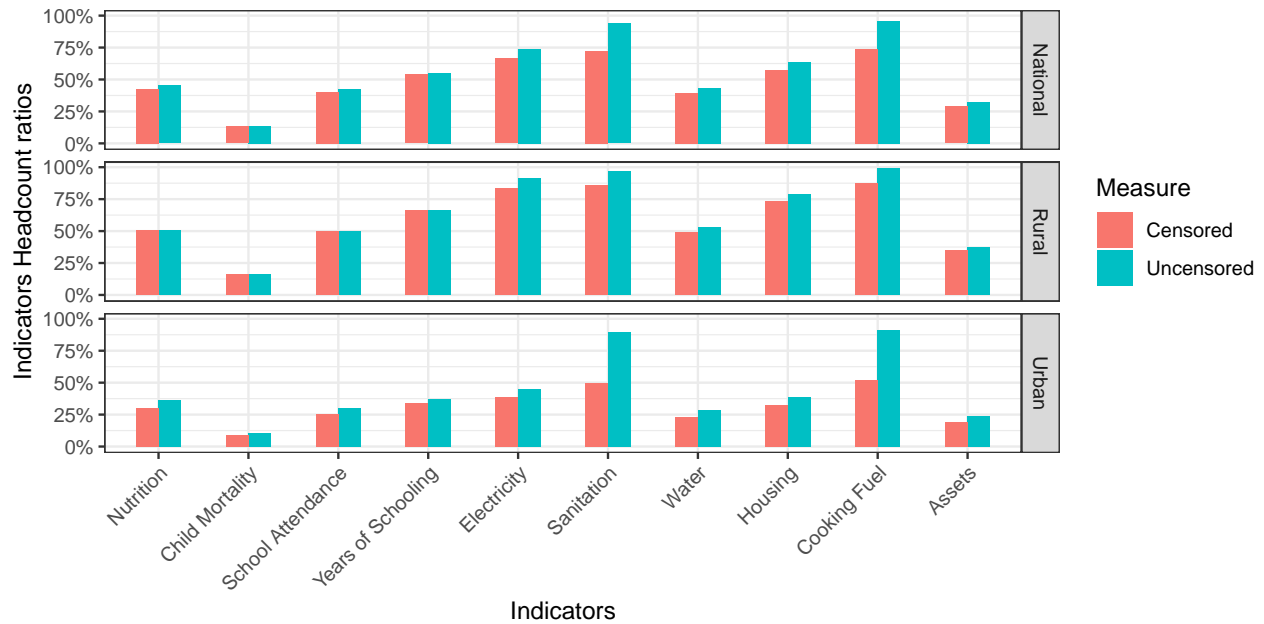


Figure 2: Indicators headcount ratios in Benin 2006

deprivation in any dimension and overall poverty. This insight could have significant implications for the design of targeted poverty reduction programs.

A concerning finding is that approximately 50% of the rural population lives with a child who is either undernourished and/or not attending school. Additionally, access to essential services like water, electricity, and adequate housing materials is notably more precarious in rural areas. Finally, educational attainment levels are significantly lower in rural compared to urban regions.

Figure 3 presents a comparative analysis of poverty across national, rural, and urban populations concerning the contribution of each indicator. The figure includes two bar plots. The left-hand panel displays the absolute contribution of each indicator to the MPI value for each subgroup. The height of each colored bar represents the absolute contribution, and the sum of all bars within a subgroup equals the total MPI value for that group. The right-hand panel illustrates the percentage contribution of each indicator to the overall MPI for each subgroup. The height of each bar represents the percentage contribution, and the sum of all bars within a subgroup equals 100%. This dual representation allows for a direct visual comparison of the composition of poverty across different population subgroups, revealing the relative importance of each deprivation dimension within each context.

```
# Filter contributions estimates
# Absolute contributions
plt_data.actb <- as.data.frame(estimate.04$lframe) %>%
  filter(measure == "actb") %>%
  # Order indicators by dimensions conveniently to plot
  # we want to avoid alphabetical order in the plot and group by dimension
  mutate(indicator = factor(stringi::stri_replace_all_regex(
    indicator, pattern = unlist(indicators),
    replacement = indicators_names, vectorize = F), levels = indicators_names)) %>%
  # Replace the subgroup names by their complete names
  mutate(subg = factor(stringi::stri_replace_all_regex(
    subg, pattern = c("nat", "rural", "urban"),
    replacement = subg_names, vectorize = F), levels = subg_names))
```

```

# Percentage contributions
plt_data.pctb <- as.data.frame(estimate.04$lframe) %>%
  filter(measure == "pctb") %>%
  # Order indicators by dimensions conveniently to plot
  # we want to avoid alphabetical order in the plot and group by dimension
  mutate(indicator = factor(stringi::stri_replace_all_regex(
    indicator, pattern = unlist(indicators),
    replacement = indicators_names, vectorize = F), levels = indicators_names)) %>%
  # Replace the subgroup names by their complete names
  mutate(subg = factor(stringi::stri_replace_all_regex(
    subg, pattern = c("nat", "rural", "urban"),
    replacement = subg_names, vectorize = F), levels = subg_names))

# Define palettes by indicators (different colors for each dimension)
palettes <- c("#A50026FF", "#D73027FF",
              "#FFFE55FF", "#FFF7BCFF",
              "#B9DDF1FF", "#94C1E0FF", "#75A6CBFF",
              "#5889B6FF", "#42779EFF", "#2A5783FF")

```

```

# Plot Absolute contribution
plt.actb <- ggplot(plt_data.actb,
                  aes(x = subg, y = b, fill = indicator)) +
  geom_bar(stat = "identity", width = 0.5) +
  # Axis Labels
  labs(y = "Contribution to MPI value", fill = "Indicators", x = "Subgroups") +
  # White background
  theme_bw() +
  # Remove legend
  theme(legend.position = "none") +
  # Colour palettes of each indicator
  scale_fill_manual(values = palettes)

# Plot Percentage contribution
plt.pctb <- ggplot(plt_data.pctb,
                  aes(x = subg, y = b, fill = indicator)) +
  geom_bar(stat = "identity", width = 0.5) +
  # Contributions as percentage
  scale_y_continuous(labels = scales::percent) +
  # Axis labels
  labs(y = "Percentage Contribution to the MPI", fill = "Indicators", x = "Subgroups") +
  # White background
  theme_bw() +
  # Legend position
  theme(legend.position = "right") +
  # Colour palettes of each indicator
  scale_fill_manual(values = palettes)

# Show plot
gridExtra::grid.arrange(plt.actb, plt.pctb, ncol = 2,
                        widths = c(0.40, 0.55),
                        heights = c(1))

```

! Users should be aware that the number of estimates generated can increase significantly depending on the

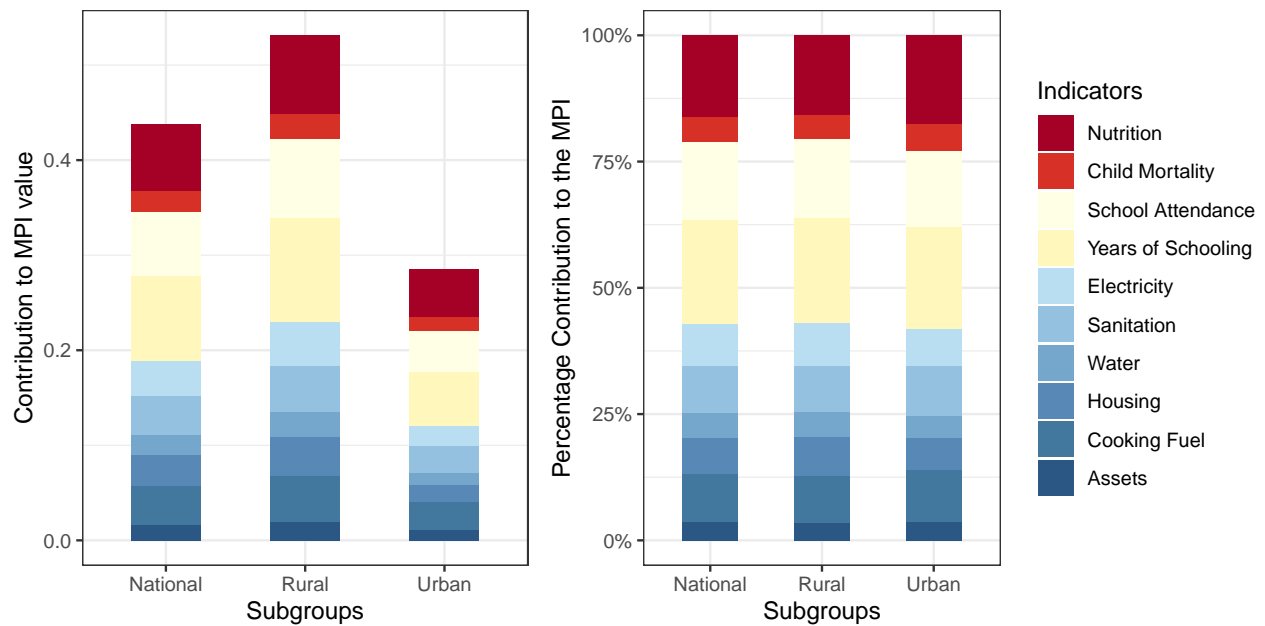


Figure 3: Indicators contributions to the MPI in Benin 2006

specific measures requested and the number of population subgroups analyzed. For example, a total of 43 measures can be calculated in a ten-indicator MPI, and this number grows proportionally with the number of population subgroups and poverty cut-offs. To optimize processing time, it is crucial to carefully select the desired measures within the `mpitb.est` arguments.

## References

- Alkire, Sabina, and James Foster. 2011. “Counting and Multidimensional Poverty Measurement.” *Journal of Public Economics* 95 (7-8): 476–87. <https://doi.org/10.1016/j.jpubeco.2010.11.006>.
- Alkire, Sabina, James Foster, Suman Seth, Maria Emma Santos, José Manuel Roche, and Paola Ballon. 2015. “Multidimensional Poverty Measurement and Analysis.” <https://doi.org/10.1093/acprof:oso/9780199689491.001.0001>.
- Alkire, Sabina, Usha Kanagaratnam, Ricardo Nogales, and Nicolai Suppa. 2022. “Revising the Global Multidimensional Poverty Index: Empirical Insights and Robustness.” *Review of Income and Wealth* 68 (S2). <https://doi.org/10.1111/roiw.12573>.
- Alkire, Sabina, and Maria Emma Santos. 2014. “Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index.” *World Development* 59 (July): 251–74. <https://doi.org/10.1016/j.worlddev.2014.01.026>.
- Lumley, Thomas. 2024. *survey: Analysis of Complex Survey Samples*. <https://CRAN.R-project.org/package=survey>.
- Suppa, Nicolai. 2023. “mpitb: a toolbox for multidimensional poverty indices.” *The Stata Journal* 23 (3): 625–57. <https://doi.org/10.1177/1536867x231195286>.